Unit 3 Project Assessment

Mark Schulist

2022-10-31

Do the majority of American adults believe that, for people convicted of murder, the death penalty is a better punishment than life imprisonment without the possibility of parole?

Study Design

The observational units in this study are the American adults' responses. There were 537 American adults that participated in this study. These adults are a simple random sample of American adults. The variable for these data is the punishment that the adults support (either death penalty or life imprisonment). This variable is qualitative.

Parameter and Hypotheses

We are interested in finding the long term probability of an American adult choosing the death penalty over life imprisonment.

H_0 = If random chance governs American adults' responses, the probability of them choosing the death penalty is equal to 0.5.

H_a = Adults prefer life imprisonment to the death penalty, thus the probability of choosing the death penalty is less than 0.5.

$$H_0: \ p = 0.5$$
$$H_a: \ p < 0.5$$

Exploring the Data

The sample size for the study is 537 as there are 537 adults surveyed. The statistic for our study is $0.47 \approx \frac{252}{537}$.

Drawing Inferences

To conduct a simulation to find the null distribution, we will use a coin with a p(heads) = 0.5. We will flip the coin 537 times as there are 537 adults surveyed. One flip of the coin represents one adult's response. We will record the number of heads for the 537 trials we will run. We will run this sequence of trials (counting the number of heads for 537 coin flips) many (10,000 times) times to get an accurate representation of the null distribution. After completing this large number of repetitions, we will find percentage of times that we got less than or equal to $0.47 \cdot 537 \approx 252$ heads in our null distribution.

Conducting a Simulation

We will conduct a simulation to get the null distribution for this study.

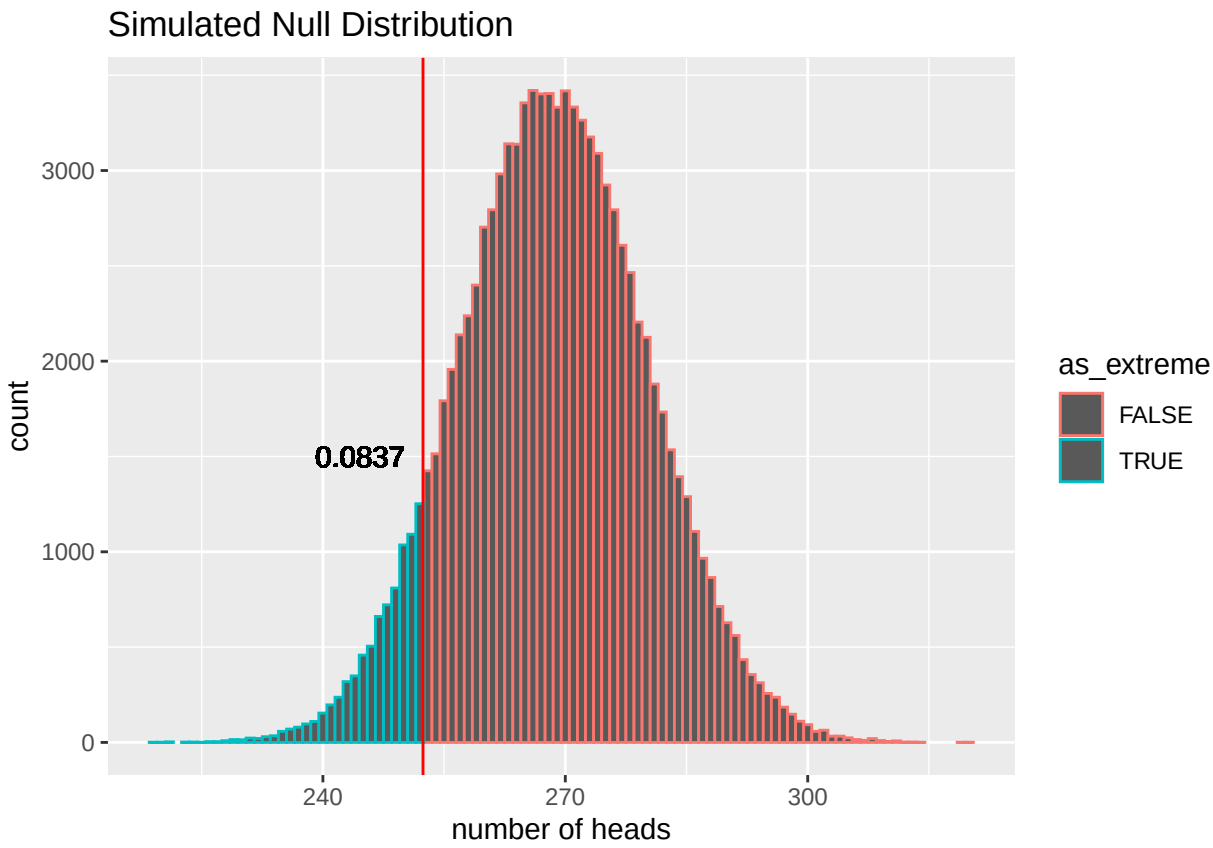
```r
set.seed(18) # for reproducible results
p <- 0.5 # prob of getting heads
n <- 537 # number of adults (number of tosses)
reps <- 100000

observed_prop <- 0.47
observed <- observed_prop * n
phat <- observed / n

sim <- rbinom(reps, n, p) %>% as_tibble() %>% rename(heads = 1) %>% # simulation
  mutate(as_extreme = (heads <= observed))

p_value = sum(sim$as_extreme == T) / reps

ggplot(sim, aes(heads)) +
  geom_bar(aes(color = as_extreme), width = 1) +
  geom_vline(xintercept = observed, color = "red") +
  geom_text(aes(observed, 1500, label = round(p_value, 4), hjust = 1.2)) +
  labs(title = "Simulated Null Distribution") +
  xlab("number of heads") +
  ylab("count")
```



From the simulation, we found that the p-value of our data is 0.08372. This p-value means if the adults randomly chose between life imprisonment and the death penalty, 8.372 times out of 100, we would get results at least as extreme as our results.

Now we are going to calculate the standardized statistic of our null distribution.

```
mean = mean(sim$heads)
SD = sd(sim$heads)
z = (observed - mean) / SD # standardized statistic
```

mean = 268.5115 responses

standard deviation = 11.6157 responses

standardized statistic = -1.3879

The standardized statistic tells us that our observed statistic ($\hat{p}$) is 1.3879 standard deviations below the mean of the null distribution (which in this case is 268.5115 responses). $\hat{p}$ is below the mean because the z-score is negative.

Theory-based Approach

Previously, we used a simulation to find the likelihood of seeing our data given the respondents had no preference between the death penalty and life imprisonment (p = 0.5). Now, we will use a theory-based approach to find this likelihood, and hopefully it will be very similar to the simulated approach.

We will first use the central limit theorem (CLT) to determine the standard deviation of the null distribution. We can assume that this null distribution will follow a normal distribution because we have at least 10 heads and 10 tails in our study.

$$\text{Standard Deviation} = \sqrt{\frac{p(1-p)}{n}}$$

$$n = \text{sample size},\ p = \text{long-run probability}$$

```
CLT_SD = sqrt(p * (1 - p) / n) * n
```

Central Limit Theorem Standard Deviation = 11.5866 responses

The mean of our theoretical null distribution is $\frac{537}{2} = 268.5$ because the probability of choosing prison or death is 0.5 when the adults are randomly choosing (don't have preference).

Now, using the theoretical standard deviation and mean, we will find the standardized statistic and p-value.

```
norm_z = (observed - n * p) / CLT_SD
norm_p_value = pnorm(q = norm_z, lower.tail = T)
```

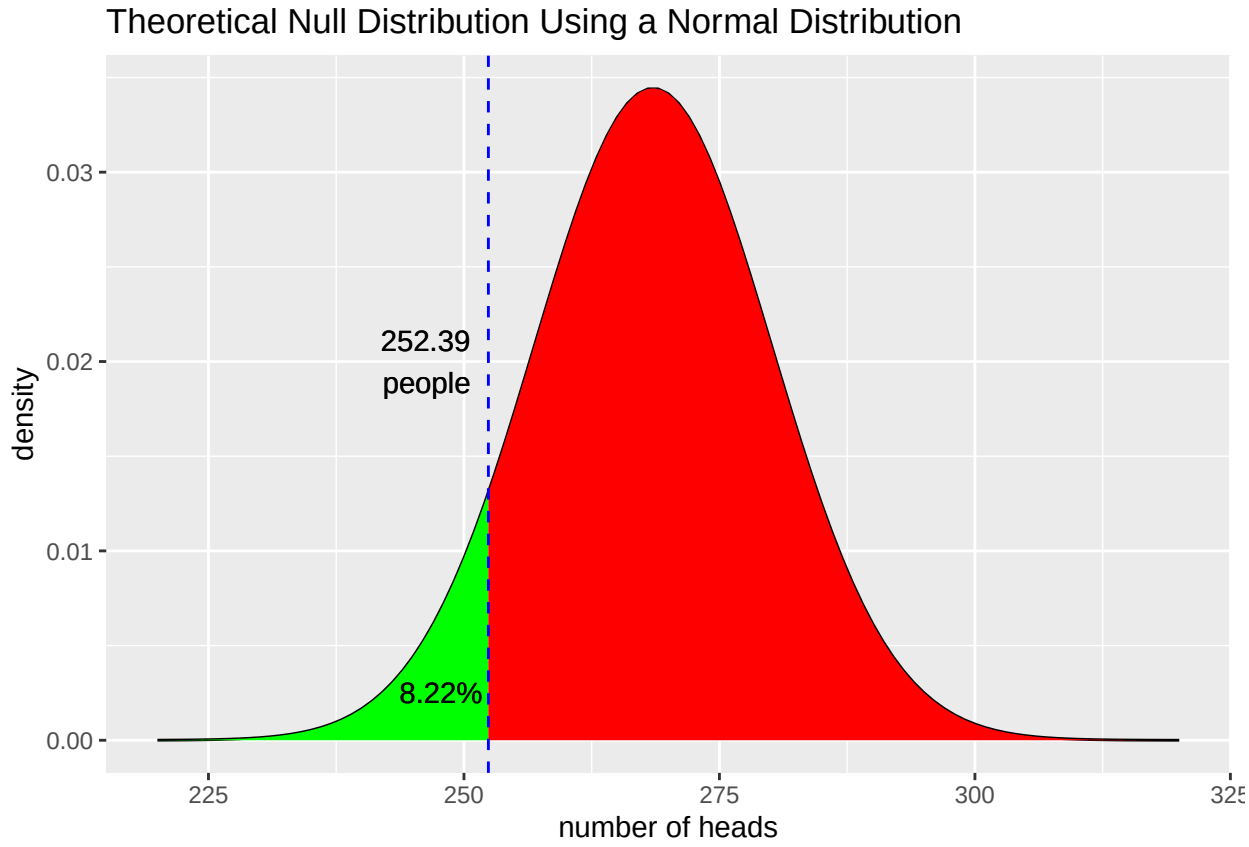
theoretical z-score = -1.3904

theoretical p-value = 0.0822

The theoretical and simulated p-values (0.0837) are extremely close to each other and, due to the large sample size, both can be used.

We can graph this standard distribution to show that it is very similar (in shape among all other aspects) to the simulated null distribution we found.

```
ggplot(data.frame(x = c(220, 320)), aes(x)) +
  stat_function(fun = dnorm, args = list(mean = n/2, sd = CLT_SD)) +
  geom_area(stat = "function", fun = dnorm, args = list(mean = n/2, sd = CLT_SD),
            fill = "green", xlim = c(220, observed)) +
  geom_area(stat = "function", fun = dnorm, args = list(mean = n/2, sd = CLT_SD),
            fill = "red", xlim = c(observed, 320)) +
  geom_vline(xintercept = observed, color = "blue", linetype = "dashed") +
  geom_text(aes(observed, 0.02, label = paste0(round(observed, 2), "\npeople"),
                hjust = 1.2)) +
  geom_text(aes(observed, 0.0025, label = paste0(round(norm_p_value * 100, 3), "%"),
                hjust = 1.07)) +
  ylab("density") +
  xlab("number of heads") +
  labs(title = "Theoretical Null Distribution Using a Normal Distribution")
```



Conclusions

Based on our p-value of 0.0837 and our standardized statistic of -1.3879, our data is not statistically significant. Our p-value is greater than 0.05 which means that there is a small, but still relevant, possibility that we could see our data from random chance. This means that we fail to reject our null hypothesis. It is not clear that the death penalty no longer has majority support: it is very possible that more than 50% of the population supports the death penalty.

I would like to know if an updated survey has come out as this study was conducted in 2006 as I'm fairly certain that public opinions respondents changed since then. It would also be nice to know their sampling method as it doesn't explicitly say how they chose their respondents.

Here's my new research question:

Do college students believe that the death penalty or life imprisonment, without the possibility of parole, is a better punishment for people convicted of murder?